network comparison

introduction to network analysis (ina)

Lovro Šubelj University of Ljubljana spring 2024/25

comparison overview

- network *comparison by isomorphism* is *NP problem* (binary)
- exact comparison requires exponentially many properties
- comparison by *graph edit distance* (add/remove nodes/edges)
- comparison by network fragments [MIK⁺04, Prž07, ARS15]
- comparison by *network distances* [SCDG⁺17, BB19]
- direct comparison of *individual metrics* [WS98, BA99, New02]
- statistical comparison over *multiple metrics* [ŠFB14, ŠBB⁺15]

comparison *metrics*

CONSISTENCY OF CITATION AND COLLABORATION TOPOLOGY OF BIBLIOGRAPHIC DATABASES

Šubelj, L., Fiala, D. & Bajec, M. Network-based statistical comparison of citation topology of bibliographic databases. Scientific Reports 4, 6496 (2014). Šubelj, L., Bajec, M., Boshkoska, B., Kastrin, A. & Levnajić, Z. Quantifying the consistency of scientific databases. PLoS ONE 10(5), e0127390 (2015).

Corresponding author: lovro.subelj@fri.uni-lj.si

NETWORKS OF BIBLIOGRAPHIC DATABASES

Charlies and collaboration networks extension from bibliographic databases. Those ext (Moly dis Comparis Educacionagos V and de Sacione and 2014 (1976) papers); (ATS) du Amenican Physical Society publications and 2014 (2018) papers); (MBL Ph Andread Carriel Carlies on pace assess publications and 2014 (2018) papers); (MBL Ph and Dath 2 - Comparis Found Bibliography and 2014 (2214) papers); (MBL Ph and Dath 2 - Comparis Found Bibliography and 2014 (2214) papers); (MBL Ph and Dath 2 - Comparis Found Bibliography and 2014 (2214) papers); (MBL Ph and Dath 2 - Comparis Found Bibliography and 2014 (2214) papers); (MBL Ph and Dath 2 - Comparis Found Bibliography and 2014 (2214) papers); (MBL Phan 2 - Comparis Found Bibliography and 2014 (2014) papers); (MBL Phan 2 - Comparison Phan

NETWORK COMPARISON METHODOLOGY

Methodogg of network-based statistical comparison of Milling-raphic databases. Natorsko representation by Bolleraphic databases and compared from 2011 graph statistics. We compare externally stadenticed statistics resoluted that measure the commany of each database with the net. Statistically application incommission in individual attatistic are revealed by independent Stadent - Hone. We select a subset of natistics where pairwise independences or withful subset of the statistics where pairwise independences resulting and the statistics where and the databases with no significant differences an avecaded by Nameraj power low to its of the databases with no significant differences an avecaded by Nameraj power low to its of the databases with no significant differences an avecaded by Nameraj power low to its of the databases.



COMPARISON OF BIBLIOGRAPHIC NETWORKS

Statistical comparison of Bibling raphic databases through statistics of networks. Panel (A) shows the critical difference diagram of Nemary into far paper cainton networks P--2, panel (B) for ambre citation networks A--A and panel (C) for ambre collenation networks A--A (so additional ambre rame disruptionies has been rank). The critical diagrams illustrate for overall ranking of the database, where flow connected by a tack has been so natistically significant increasiones at P-value = 0.1.



PROFILE OF PAPER CITATION NETWORKS

Bioithtakas, diagrams, gloss of paper datation networks extracted from Millergraphk databases based (k-4) jow (k) million this in eight) the falls bloss risk down planes, where the annexes the distribution of the links and the gamma of diagrams are proportional is the mather of abades with zero and the links and the gamma of diagrams are proportional is the mather of abades with zero and the links and the gamma of the distribution of the dis



COMPARISON OF PAPER CITATION NETWORKS

Statistical encouperious of Holling-papele databases through aduations of paper clastine networks. Proved (4-r) place sensitivation datations in the role into disconting conject, which the aduation of the properties of the provide sensitivation of adaptased to the state of the place of the properties of the place of





- statistical comparison of N networks over K metrics
- 1. x_{ij} is value of *j*th metric for *i*th network and \tilde{x}_{ij} its residual

$$\widetilde{\mathbf{x}}_{ij} = \frac{\mathbf{x}_{ij} - \widetilde{\mu}_{ij}}{\widetilde{\sigma}_{ij}\sqrt{1 - \frac{1}{N}}} \qquad \widetilde{\mu}_{ij} = \frac{1}{N-1}\sum_{k \neq i} \mathbf{x}_{kj} \qquad \widetilde{\sigma}_{ij} = \sqrt{\frac{1}{N-2}\sum_{k \neq i} (\mathbf{x}_{kj} - \widetilde{\mu}_{ij})^2} \qquad \widetilde{\mathbf{x}}_{ij} \sim t(N-2)$$

2. R_{ij} is rank of ith network for jth independent metric

 $R_{ij} = rank \ of |\widetilde{x}_{ij}| \qquad R_{ij} \in \{1, \dots, N\}$

 R_i is mean rank of ith network over K independent metrics

$$R_i = \frac{1}{K} \sum_j R_{ij}$$
 $\frac{12K}{N(N+1)} \left(\sum_i R_i^2 - \frac{N(N+1)^2}{4} \right) \sim \chi^2(N-1)$

3. $|R_i - R_j|$ statistically significant when above critical difference $q\sqrt{\frac{N(N-1)}{6K}}$

comparison references



David Aparício, Pedro Ribeiro, and Fernando Silva. Network comparison using directed graphlets. *e-print arXiv:151101964v1*, 2015.



A.-L. Barabási and R. Albert.

Emergence of scaling in random networks. Science, 286(5439):509-512, 1999.



A.-L. Barabási.

Network Science. Cambridge University Press, Cambridge, 2016.



James P. Bagrow and Erik M. Bollt.

An information-theoretic, all-scales approach to comparing networks. Appl. Netw. Sci., 4:45, 2019.



Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.

Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition. Cambridge University Press, Cambridge, 2011.



David Easley and Jon Kleinberg.

Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, Cambridge, 2010.



Ernesto Estrada and Philip A. Knight.

A First Course in Network Theory. Oxford University Press, 2015.

comparison references



Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal

Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004.



M. E. J. Newman.

Assortative mixing in networks. Phys. Rev. Lett., 89(20):208701, 2002.



Mark E. J. Newman.

Networks. Oxford University Press, Oxford, 2nd edition, 2018.



Nataša Pržulj.

Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, 2007.



Lovro Šubelj, Marko Bajec, Biljana Mileva Boshkoska, Andrej Kastrin, and Zoran Levnajić. Quantifying the consistency of scientific databases. *PLoS ONE*, 10(5):e0127390, 2015.



Tiago A. Schieber, Laura Carpi, Albert Díaz-Guilera, Panos M. Pardalos, Cristina Masoller, and Martín G. Ravetti.

Quantification of network structural dissimilarities. Nat. Commun., 8:13928, 2017.



Lovro Šubelj, Dalibor Fiala, and Marko Bajec.

Network-based statistical comparison of citation topology of bibliographic databases. *Sci. Rep.*, 4:6496, 2014.

comparison *references*



D. J. Watts and S. H. Strogatz.

Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.